260585

CATALOGED BY ASTIA
AS AD No. ______

Technical Report No. 32-123

Classification and Evaluation of Coherent Synchronous Sampled-Data Telemetry Systems

Andrew Viterbi

474 200

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

June 15, 1961

S 61-39491

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
CONTRACT NO. NASw-6

Technical Report No. 32-123

CLASSIFICATION AND EVALUATION OF COHERENT SYNCHRONOUS SAMPLED-DATA TELEMETRY SYSTEMS

Andrew Viterbi

W. K. Victor

W. Victor, *Chief*
Communications Systems
Research Division

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA
June 15, 1961

CONTENTS

FIGURES

FIGURES (Cont'd)

CLASSIFICATION AND EVALUATION OF COHERENT SYNCHRONOUS SAMPLED-DATA TELEMETRY SYSTEMS*

Andrew Viterbi
Jet Propulsion Laboratory
Pasadena, California

Summary

This paper analyzes the various types of continuous wave and pulse modulation for the transmission of sampled data over channels perturbed by white gaussian noise. Optimal coherent synchronous detection schemes for all the different modulation methods are shown to belong to one of two general classes: linear synchronous detection and correlation detection. The figures of merit, mean-square signal-to-error ratio and bandwidth occupancy, are determined for each system and compared.

Introduction

The consideration of transmission methods for sampled data is a significant communications problem for several reasons. First, by virtue of the well-known sampling theorem[1], any signal may be presented as sampled data with no loss of information provided the sampling rate is greater than twice its highest frequency. Second, certain types of modulation, notably those involving pulses, require sampling of the signal prior to modulation. Finally, the most practical reason is that certain types of data sources are of a sampled nature; i.e., commutators are employed which sample a given source periodically. Sampled data signals also afford an inherent degree of synchronization which can be used advantageously in a communication system.

For the evaluation and comparison of various modulation and detection systems, a generic performance criterion must first be stipulated. One of the simplest to measure and to calculate is a mean-square error criterion, which will be discussed in the next section. On this basis, a wide class of coherent synchronous communication systems will then be analyzed.

Error Criteria

A general communication system for sampled data signals is shown in Fig. 1. The channel interference is assumed to be white gaussian noise. Clearly, the value of the system depends on how nearly the output signal (within a scale factor) matches the transmitted signal. Path loss and transmitter and receiver gains will, of course, vary the output signal amplitude. Hence, for a comparison with the output signal, the data signal must be amplified or attenuated by an appropriate factor K. Thus, the error in the nth sample will be $\epsilon_n = s_n - o_n$, where s_n would be the nth output sample in the absence of noise, and o_n is the actual nth output sample. It may be assumed that both the data signal and the output signal levels have been adjusted so that their means are zero. Then the error mean will also be zero.

There are several measures of how nearly the output resembles the data. The most common is the mean-square error which for stationary signal and noise is given by

$$\overline{\epsilon_n^2} = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \epsilon_n^2$$

A simple method of measuring this is shown in Fig. 1. A criterion which is more simply measured and generally equally valid is the mean-absolute error

$$\overline{|\epsilon|} = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} |\epsilon_n|$$

Although the mean-absolute error is precisely the same as the mean-square error when both the signal and noise are gaussian, it is not readily calculated when one of the two is not gaussian. Likewise, several other error criteria are equivalent to the mean-square error for gaussian distributions[2] but are hopeless to determine in general. Thus, for the sake of analytical feasibility, only the mean-square error will be considered. Of course, the magnitude of this parameter will depend on the data signal magnitude. For this reason it is necessary to normalize $\overline{\epsilon_n^2}$ by some parameter of the signal. The most significant such parameter is the data signal power or mean-square signal

$$\overline{s_n^2} = \lim_{N \to \infty} \frac{1}{N} \sum (s_n)^2$$

It is evident that both $\overline{s_n^2}$ and $\overline{\epsilon_n^2}$ have the dimension of power. For the sake of similarity to the well-known signal-to-noise ratio, the ratio $\overline{s_n^2}/\overline{\epsilon_n^2}$, or mean-square signal-to-error ratio, will be used here.

Classification of Telemetry Systems

A number of modulation methods of the continuous wave and pulse variety are in common use. Certain types (such as amplitude modulation) are basically for transmission of continuous signals rather than sampled data. However, even these have an equivalent form for sampled data. The following is a list of the forms of modulation which are generally accepted[3] and their sampled data equivalents:

1. Amplitude modulation, or pulse amplitude modulation (PAM)
2. Phase modulation, or phase shift keying (PSK)
3. Frequency modulation, or frequency shift keying (FSK)
4. Pulse duration modulation (PDM)
5. Pulse position modulation (PPM)
6. Pulse code modulation (PCM)

*This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract NASw-6, sponsored by the National Aeronautics and Space Administration.

In this work, we shall be interested in coherent synchronous detection of the various forms of modulation; that is, the detection process will involve locally generated signals which are coherent to the transmitted carrier and synchronous with the sampled data rate. In this context, it will be shown that the six types of modulation may be detected coherently and synchronously by means of one of the following detection techniques:

1. Linear synchronous detection
2. Correlation detection, or matched filtering

The next two sections will treat the detection of the six forms of modulation by these two methods.

Linear Synchronous Detection

In this section it will be shown that pulse amplitude modulation (PAM), pulse duration modulation (PDM), and phase shift keying (PSK), can all be detected by a pulsed integrator synchronous with the sample transmission time and that the output signal-to-mean-square-error ratio is a linear function of the channel signal-to-noise ratio.

Pulse amplitude modulation consists simply of extending the sample amplitude to last over the allotted transmission time of T seconds and multiplying the carrier by this waveform. This is shown in Fig. 2a. In pulse duration modulation, a pulse is generated whose width is proportional to the amplitude of the sample. For this purpose, the sample must be amplitude-limited, say between $-A$ and $+A$. Then a sample of amplitude x_n (where $-A < x_n < A$) will produce a pulse of duration $[1 + (x_n/A)](T/2)$ seconds. In order to keep the transmitted power a constant, the pulse waveform is made to alternate between 1 and -1 rather than between 1 and 0. The waveform is then used to multiply the carrier (Fig. 2b). In phase shift keying, the amplitude of the sinusoidal carrier is not varied, but rather its phase is varied from $-\pi/2$ to $\pi/2$, depending on the amplitude of the sample. Again, the data is assumed to be amplitude-limited between $+A$ and $-A$, and the phase of the carrier over a given sample transmission time is $\sin^{-1} x_n/A$ when the sample is x_n (where $-A \leq x_n < A$). Since the phase varies between $-\pi/2$ and $\pi/2$ it is, therefore, unambiguous. The phase waveform is shown in Fig. 2c. Both PDM and PSK have the advantage that they maintain a constant transmitted power, but in return they require a limited data amplitude. In PAM, the power varies from one sample transmission period to the next. A stringent limit need not be placed on the data amplitude, but peak power limitations are, nevertheless, present in the transmitter.

Mean-square signal-to-error ratio

The remarkable aspect of these three different modulated signals is that they can all be demodulated by the same synchronous detector; namely, a pulsed integrator (Fig. 3). It should be noted, first of all, from Fig. 2 that the power transmitted for PDM and PSK is KS watts, where S is the received power and $1/K$ the channel attenuation. Furthermore, if the data signal has a flat distribution between A and $-A$, as will be assumed henceforth, the transmitted power for PAM will also be KS watts. This follows from the fact that the power P in the transmitted signal envelope is

$$\int_{-\infty}^{\infty} x^2\, p(x)\, dx = \frac{1}{2(6KS)^{1/2}} \int_{-(6KS)^{1/2}}^{(6KS)^{1/2}} x^2\, dx = 2KS$$

where $p(x)$, the probability density of the envelope of the transmitted signal, is $1/2(6KS)^{1/2}$ between $-(6KS)^{1/2}$ and $(6KS)^{1/2}$ and is zero elsewhere. The envelope power must be divided by 2 to determine the power in the modulated sinusoid. It should be noted, however, that the power during a given sample transmission period of T seconds will vary from 0 to $3S$ watts; hence, the peak power is three times that for PDM or PSK.

Thus, the average signal power into the synchronous detector will be S. The first stage of the demodulator consists of a multiplier that is phase coherent with the received carrier, which shifts the spectrum to low frequencies. In the case of PAM (Fig. 3a), the multiplier output over the nth sample period is

$$\left(\frac{3S}{A}\right)^{1/2} x_n(1 + \cos 2\omega_0 t)$$

The double-frequency term may be neglected since it will be discarded by the integrator, as will be shown. For PDM, the low-frequency output of the multiplier is the same as the modulating signal at the transmitter (Fig. 2b) except for the gain factor of $S^{1/2}$ (Fig. 3b). For PSK, the output of the multiplier during the nth sample period is

$$(2S)^{1/2} \sin\left(\omega_0 t + \sin^{-1}\frac{x_n}{A}\right) \cdot 2^{1/2} \cos \omega_0 t$$

$$= S^{1/2}\frac{x_n}{A} + S^{1/2} \sin\left(2\omega_0 t + \sin^{-1}\frac{x_n}{A}\right)$$

Thus, the low-frequency component is $S^{1/2}x_n/A$ (Fig. 3c). The second stage of the synchronous detector is a pulsed integrator which during each sample integrates for T seconds and is followed by an amplifier or attenuator with gain $1/T$. For the case of PAM, the low-frequency component produces an output at time T of $(3S)^{1/2}x_n/A$, while the double-frequency term produces

$$(3S)^{1/2}\frac{x_n}{A}\frac{\sin 2\omega_0 t}{2\omega_0}$$

at time T. If T is chosen such that $T = \pi k/2\omega_0$, where k is an integer (that is, if the data rate is a certain multiple of the carrier frequency) then the double-frequency term is 0 and may properly be neglected for all three types of modulation. For PDM, the integrator will produce a ramp of slope $S^{1/2}/T$ for $[1 + (x_n/A)]T/2$ seconds and a ramp of slope $-S^{1/2}/T$ for the remaining $[1 - (x_n/A)]T/2$ seconds (Fig. 3b). Thus, the net output amplitude at the end of the sample transmission period is $S^{1/2}x_n/A$. For PSK, the integrator will produce a ramp of slope $S^{1/2}x_n/TA$ for T seconds, and thus, an amplitude of $S^{1/2}x_n/A$ at the end of the integrating period, the same as for PDM. Since the signal was assumed to have a flat distribution between $-A$ and A, the output signal power or mean-square signal at the end of the sample transmission time will be for PDM and PSK:

$$\overline{s_n^2} = \int_{-A}^{A} S\left(\frac{x_n}{A}\right)^2 p(x_n)\, dx_n = \frac{S}{A}\int_{-A}^{A}\frac{x_n^2}{2A}\, dx_n = \frac{S}{3}$$

For PAM, since the output signal is three times as large, $\overline{s_n^2} = S$

Clearly, the output error will be the same in all cases since the same detector is used. It is assumed that the received noise $N(t)$ is white gaussian noise and has spectral density $N/2B$, where N is the noise power measured at the output of a bandpass filter of bandwidth B cps. The noise output of the multiplier will be $2^{1/2}N(t)\sin\omega_0 t$. Then the variance or mean-square error at the output of the synchronous detector at the end of the sample transmission period is

$$\overline{\epsilon_n^2} = \sigma^2 = \frac{1}{T^2}\int_0^T\int_0^T \overline{2N(t)\,N(u)}\,\sin\omega_0 t\,\sin\omega_0 u\; dt\, du$$

$$= \frac{1}{T}\,(N/2B)$$

Hence, in the case of PDM and PSK, the ratio of output signal power-to-mean-square error is

$$\overline{s_n^2}/\overline{\epsilon_n^2} = \frac{2\,ST}{3\,N/B}$$

For PAM, the ratio is

$$\overline{s_n^2}/\overline{\epsilon_n^2} = \frac{2\,ST}{N/B}$$

It should be emphasized that the factor of three superiority of PAM over PDM and PSK is due solely to the lack of a constant power restriction in this case. If the restriction were placed on PAM that the energy per sample transmission period were not to exceed that for PDM and PSK (i.e., if a peak power rather than an average power restriction were used) then all three forms of modulation would yield the same result.

The mean-square signal-to-error ratio at the output of the detector is shown in Fig. 4 as a function of $ST/(N/B)$, the (received energy per sample)/(noise spectral density). It is now evident that the performance of this form of synchronous detection is a linear function of the channel parameters.

Bandwidth occupancy

A significant consideration in the evaluation of any communication system is the bandwidth which it occupies. In this treatment, the bandwidth occupancy of a channel will be defined as the minimum frequency separation required between the given channel and an adjacent channel modulated in precisely the same manner so that the adjacent channel will have no effect on the detector for the given channel. In the case of PAM with a sample transmission period T, an adjacent channel similarly modulated must be placed $1/T$ cps away, or a multiple thereof, in order that the detector for a given channel will not be influenced. This is shown by the fact that the detector at the end of the transmission period will produce an output due to the adjacent channel of

$$\int_0^T 2^{1/2}\sin\omega_0 t\cdot 2^{1/2}\sin\left[\left(\omega_0+\frac{2\pi}{T}\right)t+\phi\right] = 0$$

where ϕ is the arbitrary initial phase difference between the adjacent channel carriers. If ϕ could be made zero so that the various channel carriers were phase coherent, then the bandwidth separation of the channels could be cut in half to $1/2T$ cps.

The situation is the same for PSK since if the adjacent channel is transmitting a sample of amplitude y_n during a given transmission period, the output of the given channel detector at the end of the period will be

$$\int_0^T 2^{1/2}\sin\omega_0 t\cdot 2^{1/2}\sin\left[\left(\omega_0+\frac{2\pi}{T}\right)t+\sin^{-1}\frac{y_n}{A}+\phi\right]dt = 0$$

However, in this case the bandwidth occupancy cannot be cut in half by making $\phi = 0$, since the phase is modulated and, hence, adjacent channels can not be made phase coherent.

Circumstances are less favorable for PDM. If the adjacent channel carrier is taken ω_D radians from the given channel carrier, a sample of amplitude y_n modulating the adjacent channel will produce a waveform at the given channel detector input which is $\sin[(\omega_0+\omega_D)t+\phi]$ for the first $[1+(y_n/A)]\,T/2$ seconds of the transmission period and is $-\sin[(\omega_0+\omega_D)t+\phi]$ for the remainder. Thus, the given channel detector will produce an output at the end of the period which is

$$\frac{1}{T}\left\{\int_0^{\left(1+\frac{y_n}{A}\right)\frac{T}{2}} 2^{1/2}\sin\omega_0 t\cdot 2^{1/2}\sin[(\omega_0+\omega_D)t+\phi]\,dt - \int_{\left(1+\frac{y_n}{A}\right)\frac{T}{2}}^{T} 2^{1/2}\sin\omega_0 t\cdot 2^{1/2}\sin[(\omega_0+\omega_D)t+\phi]\,dt\right\}$$

$$\approx \frac{2\sin\left[\frac{\omega_D T}{2}\left(1+\frac{y_n}{A}\right)+\phi\right]-\sin[\omega_D T+\phi]}{\omega_D T}$$

The approximation is valid since the double frequency terms will be divided by $2\omega_0+\omega_D >> \omega_D$. Then the detector output due to the adjacent channel will be at all times less than $3/\omega_D T$, but it can not be made precisely zero because of the random nature of y_n. In order to maintain the channel cross-modulation below 1% at all times, it is necessary to make

$$\omega_D \geq \frac{300}{T}\ \frac{\text{rad}}{\text{sec}}$$

which means that the frequency separation must be greater than about $50/T$ cps. This is a serious handicap for PDM relative to PAM and PSK. These various results are summarized in Table 1.

Correlation Detection

The forms of modulation which were not treated under linear synchronous detection were pulse position modulation (PPM), frequency shift keying (FSK), and pulse code modulation (PCM).

The latter is strictly digital and needs to be considered separately in any case. PPM and FSK, however, could be used to transmit an analog sample. PPM involves varying the position or leading edge of a narrow pulse throughout the sample transmission period according to the sample amplitude. FSK varies the carrier frequency according to the sample amplitude. However, neither of these can be demodulated by a linear synchronous detector because this requires an averaging process over the transmission time. In PPM, the signal exists for only a small portion of this time; for FSK, time averaging can not be used to detect a frequency. On the other hand, the synchronous nature of sampled data can be used to advantage if the samples are quantized into L levels, as will now be shown first for PPM and then for FSK.

Quantized PPM

The block diagram for a quantized PPM modulator and demodulator is shown in Fig. 5. The sampled data is quantized into L levels (where L will be assumed even); depending upon the level of a given sample, a pulse will be generated in one of L possible positions in the sample transmission interval of duration T. Again, the data is assumed to have a flat amplitude distribution between $-A$ and A. If the amplitude of a given sample lies between $Ak/(L/2)$ and $A(k+1)/(L/2)$ (where k is a positive integer less than $L/2$), a pulse will be sent in the $(L/2+k)$th position; while if it lies between $-A(k+1)/(L/2)$ and $-Ak/(L/2)$, it will be sent as a pulse in the $(L/2-k)$th position. This is multiplied by the carrier and transmitted.

The receiver is again considered to be both phase coherent with the transmitted carrier and synchronous to the sampling period. The received signal is assumed to have power S watts. Thus, each pulse of width T/L must be of amplitude $(2SL)^{1/2}$. If the path loss is $1/K$, the transmitted power must be KS watts. The output of the first multiplier in the demodulator has a low-frequency component which is a pulse of amplitude $(SL)^{1/2}$ and of width T/L during a given sample transmission period (Fig. 5) plus a double-frequency component which is eliminated in the detection, as will be shown. This signal is fed to a bank of L correlation detectors, each of which is matched to one of the L possible pulse positions. Each detector consists of a multiplier whose other input during the given period is a pulse occurring at one of the L possible positions; this is followed by an integrator of gain L/T which integrates over the sample transmission period and is then discharged. Thus, in the absence of noise the correlator corresponding to the received pulse position will produce an output of magnitude $(SL)^{1/2}$ while all the other correlator outputs will be zero. All these outputs are sampled at the end of the sample transmission period and fed to a decision device which will select the greatest output to be the correct pulse position. This is known as a maximum likelihood detector and was first proposed by Woodward.[4]

Noise in the channel of spectral density $N/2B$ watts/cps will produce a variance at the output of each correlator at the end of the sample transmission period equal to

$$\sigma^2 = \left(\frac{L}{T}\right)^2 \int_0^{\frac{T}{L}} \int_0^{\frac{T}{L}} (N/2B)\,\delta(t-u)\,(2^{1/2}\sin\omega_0 t)(2^{1/2}\sin\omega_0 u)\,dt\,du$$

$$= \frac{L}{T}\,(N/2B)$$

since the time over which the noise is integrated is T/L. Then, the ratio of correlator output to rms error for that correlator corresponding to the correct pulse position is

$$\frac{\sqrt{SL}}{\sigma} = \left(\frac{2\,ST}{N/B}\right)^{1/2}$$

Furthermore, for white gaussian noise the outputs of the n detectors due to noise will be uncorrelated since each detector integrates over a different time period of duration T/L, and the input noise in each period is independent of the noise over any other because it is white.

Before the mean-square signal-to-error ratio $\overline{s_s^2}/\overline{\epsilon_n^2}$ can be determined, the probability of error in detecting any given sample must be calculated. The probability of detecting a given sample correctly is equal to the probability that the correlator corresponding to the position transmitted will have a greater output than all the others. Thus, for gaussian noise the probability of error, which is 1 minus the probability of correct reception, is given by

$$P_E = 1 - P(x_1 > x_2, x_3, \cdots x_L)$$

$$= 1 - \prod_{i=2}^{n} P(x_i < x_1)$$

$$= 1 - \int_{-\infty}^{\infty} \frac{e^{-\frac{(x_1-\sqrt{SL})^2}{2\sigma^2}}}{(2\pi)^{1/2}\sigma}\,dx_1 \left[\int_{-\infty}^{x_1} \frac{e^{-\frac{x_i^2}{2\sigma^2}}}{(2\pi)^{1/2}\sigma}\,dx_i\right]^{L-1}$$

where

$$\sigma^2 = \frac{L}{T}\,(N/2B)$$

The second equality holds because the various correlator outputs are independent. By proper substitutions this equation can be shown to become

$$P_E = 1 - \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{(2\pi)^{1/2}}\,dy \left[\int_{-\infty}^{y+\left(\frac{2ST}{N/B}\right)^{1/2}} \frac{e^{-\frac{z^2}{2}}}{(2\pi)^{1/2}}\,dz\right]^{L-1}$$

Thus, the error probability is a function of the (received energy per sample)/(noise spectral density). This was evaluated by means of an IBM 704 computer for $L = 4$, 8, 16, 32, and 64. This is shown in Fig. 6.

As a function of probability of error, the output mean-square signal-to-error ratio can now be determined. Since the noise outputs of all the correlators are independent, the probability of the result falling into any particular incorrect level is equal to the probability of its falling into any other incorrect level. However, the error amplitude is not independent of the signal amplitude since, for example, if a sample was sent which was in the highest positive level, only a negative error can result. Figure 7 shows the transition diagram from transmitted level to detected level. Since P_E is the probability of making any error and there are $L-1$ possible errors as well as L possible transmitted levels, the probability of any transition other than the correct one is $P_E/[L(L-1)]$. It is clear from Fig. 7 that there are $2(L-1)$ ways in which a one-level error can be made, $2(L-2)$ ways in which a two-level error can be made, down to only two ways in which an $(L-1)$-level error can be made. Positive errors and negative errors are equally likely. The sample amplitude is again evenly distributed between $-A$ and A. Thus, the probability distribution of errors is as shown in Fig. 8a.

Since the sample amplitude is evenly distributed, the error due to quantization is evenly distributed between $-A/L$ and A/L about each level. Thus, the overall error probability density is as shown in Fig. 8b.

On this basis, the mean-square error can be computed in terms of P_E:

$$\overline{\epsilon_n^2} = \int_{-\infty}^{\infty} x^2\, p(x)\, dx$$

$$= \frac{1-P_E}{3}\left(\frac{A}{L}\right)^2 + \frac{P_E}{3(L-1)}\left(\frac{A^2}{L^3}\right)$$

$$\times \sum_{i=1}^{L-1} (L-i)\left[(2i+1)^3 - (2i-1)^3\right]$$

$$= \frac{A^2\left[1 + 2P_E\,(L^2 - L)\right]}{3L^2}$$

At the same time, since the samples are evenly distributed between $-A$ and A, then

$$\overline{s_n^2} = \int_{-A}^{A} x^2\, p(x)\, dx = \int_{-A}^{A} \frac{x^2}{2A}\, dx = \frac{A^2}{3}$$

Then the mean-square signal-to-error ratio is

$$\overline{s_n^2}/\overline{\epsilon_n^2} = \frac{1 + 2\,P_E(L^2 - L)}{L^2}$$

where P_E is given as a function of $ST/N/B$ and L in Fig. 6. Combining these results $\overline{s_n^2}/\overline{\epsilon_n^2}$ is shown as a function of $ST/N/B$ for $L = 4, 8, 16, 32$, and 64 in Fig. 4.

Quantized FSK

The significant feature of the PPM system just described is the orthogonality of its transmitted signals. That is, the pulse signal $x_i(t)$ representing any given level over a sample transmission period is orthogonal to that representing any other level $x_j(t)$; i.e.,

$$\int_0^T x_i(t)\, x_j(t)\, dt = 0$$

This orthogonality can be achieved in a multitude of ways. For example, two sinusoidal signals differing in frequency by $1/T$ cps will be orthogonal over an interval of T seconds since

$$\int_0^T \sin \omega t \,\sin\left[\left(\omega + \frac{2\pi}{T}\right)t + \phi\right] dt = 0$$

This suggests the possibility of encoding the various quantization levels into a set of sinusoids spaced $1/T$ cps apart from one another in frequency and of duration T seconds. This system, which may be called quantized frequency shift keying, is shown in Fig. 9. Each sample is quantized and, depending on its level, one of the L stored (or locally generated) sinusoids is transmitted over the sample transmission period T. The demodulator consists of a bank of correlation detectors each of which multiplies the received signal by one of the L frequencies and then integrates the product synchronously for T seconds and attenuates it by $1/T$. In the absence of noise, the correlator corresponding to the received signal produces an output at time T which is

$$\frac{1}{T}\int_0^T 2S^{\frac{1}{2}} \sin^2\left(\omega_0 + \frac{2\pi n}{T}\right)dt = S^{\frac{1}{2}}\left[1 - \frac{\sin 2(\omega_0 T + 2\pi n)}{2(\omega_0 T + 2\pi n)}\right]$$

If ω_0 is a multiple of $\pi/2T$, then the output is simply $S^{\frac{1}{2}}$. The output of all the correlators will be zero because of the orthogonality. It should be evident at this point that the operation and evaluation of this FSK system is identical to that of the PPM system described above and that both the error probability and the $\overline{s_n^2}/\overline{\epsilon_n^2}$ ratio will be the same.

Bandwidth occupancy

It will now be shown that both the quantized PPM and FSK systems utilizing correlation detection utilize a bandwidth of L/T cps, where L is the number of quantization levels. It is clear from the previous discussion that this holds for FSK since FSK channels can be placed on both sides of the given channel, provided the frequencies of the adjacent channels are also spaced $1/T$ cps apart and the highest and lowest frequencies of the adjacent channels are placed $1/T$ cps from the lowest and highest frequency of the given channel.

For quantized PPM, if the carrier of the adjacent channel is placed L/T cps from the given channel, then the correlator for the given channel which corresponds to the pulse position of the adjacent channel will produce an output

$$\int_0^{\frac{T}{L}} 2 \sin \omega_0 t \,\sin\left[\left(\omega_0 + \frac{2\pi L}{T}\right)t + \phi\right] dt = 0$$

and, hence, no interaction occurs.

Finally, it should be noted that if the FSK frequencies can be made phase coherent or if the carriers for adjacent PPM channels can be made phase coherent, the same results can be obtained, but now the utilized bandwidth is cut in half and becomes $L/2T$ cps. This can be seen by changing the pertinent equations so that $\phi = 0$, and the second sinusoid is spaced half as far in frequency from the first. It will be seen that the integral of the product over the given time interval is still zero. These results are summarized in Table 1.

Pulse Code Modulation

Nonredundant codes

A pulse code modulation system is defined as one in which the samples are quantized and binary codes are sent to represent the various data levels. The simplest such system transmits the binary equivalent of the numeral value for the given data level by transmitting the pure carrier to represent a *zero* and modulating the carrier by π radians to represent a *one* (Fig. 10). Thus, an L-level sample will be represented by a binary code of length $\log_2 L$ bits. The detector is a synchronous integrator which operates on each bit at a time for a period of $T/\log_2 L$ seconds and then decides whether a *zero* or a *one* was sent during that time. As described, this system utilizes correlation detection on a portion of the received sample code and should, therefore, be inferior to a system which detects the entire sample signal at once.

To determine the $\overline{s_n^2}/\overline{\epsilon_n^2}$ ratio, the probability of error per bit must be determined first. The signal in the absence of noise at the end of the bit period will be $\pm S^{1/2}$, the sign depending upon whether a *zero* or a *one* was sent. The variance is

$$\sigma^2 = \left(\frac{\log_2 L}{T}\right)^2 \int_0^{T/\log_2 L}\int (N/2B)\, \delta(t-u)\, (\sqrt{2}\sin\omega_0 t)$$

$$\times (\sqrt{2}\sin\omega_0 u)\, dt\, du$$

$$= \frac{\log_2 L}{T}\,(N/2B)$$

Thus, the signal-to-rms error at time T when a *one* was sent is $[2ST/\log_2 L(N/B)]^{1/2}$. The probability of bit error when a *one* was sent is then the probability that the noise contribution at time T is less than $-S^{1/2}$ or

$$P_B = \int_{-\infty}^{-\left[\frac{2\,ST}{\log_2 L(N/B)}\right]^{1/2}} \frac{e^{-\frac{x^2}{2}}}{(2\pi)^{1/2}}\,dx$$

By symmetry, the error probability is the same when a *zero* was sent. This quantity is plotted as a function of $ST/\log_2 L(N/B)$ in Fig. 11. The mean-square error can be determined from the probability of error, as was done in a previous section (Correlation Detection). For example, for $L = 4$ levels, the transition diagram is shown in Fig. 12. From this diagram, the error distribution can be determined (Fig. 13): and from Fig. 13, the mean-square error can be determined as

$$\overline{\epsilon_n^2} = \int x^2\, p(x)\, dx = \frac{A^2}{3(4)^2}\left[P(0) + 2\sum_{i=1}^{3} P(i)\,(12i^2+1)\right]$$

$$= \left(\frac{1+60\,P_B}{48}\right)A^2$$

where $P(i)$ is the probability of an i-level error. The mean-square signal is as before

$$\overline{s_n^2} = A^2/3$$

Then

$$\overline{s_n^2}/\overline{\epsilon_n^2} = \frac{16}{1+60\,P_B} \qquad (L=4)$$

Similarly, for $L = 8$, 16, 32, and 64 the mean-square error can be determined to be

$$\overline{s_n^2}/\overline{\epsilon_n^2} = \frac{64}{1+252\,P_B} \qquad (L=8)$$

$$\overline{s_n^2}/\overline{\epsilon_n^2} = \frac{256}{1+1020\,P_B} \qquad (L=16)$$

$$\overline{s_n^2}/\overline{\epsilon_n^2} = \frac{1024}{1+4092\,P_B} \qquad (L=32)$$

$$\overline{s_n^2}/\overline{\epsilon_n^2} = \frac{4096}{1+16{,}380\,P_B} \qquad (L=64)$$

Combination of these results with those of Fig. 11 of P_B as a function of $ST/(N/B)$ yields $\overline{s_n^2}/\overline{\epsilon_n^2}$ as a function of $ST/(N/B)$, which is plotted in Fig. 4 and is there compared with the other forms of modulation.

The bandwidth occupancy for nonredundant PCM is inversely proportional to the code bit period and hence is $(\log_2 L)/T$ cps.

Orthogonal codes

It can be shown[5] that a set of orthogonal signals can be generated with binary codes. This suggests the possibility of correlation detection for PCM; that is, each level will correspond to one binary code word which is orthogonal to all the rest. The receiver will consist of L correlation detectors, one for each binary code word[5]. Such a PCM system is equivalent to the quantized PPM and FSK systems, discussed previously, and the results for both $\overline{s_n^2}/\overline{\epsilon_n^2}$ and bandwidth occupancy are the same.

Conclusions

Figure 4 and Table 1 present the main results of this paper, and from these the pertinent conclusions can be drawn. It is clear that for reasonably high mean-square signal-to-error ratios, those modulation systems which are demodulated by correlation detection are significantly superior to those demodulated by linear synchronous detection since the former improve exponentially with increasing $ST/(N/B)$, while the latter improve only linearly.

The saturation of the curves for the correlation detection systems of Fig. 4 is due to the quantization of the samples. No matter how low the channel noise may become, the quantization noise will always be present and constant for a fixed number of levels.

Nonredundant PCM, which requires roughly $1/L$th as much demodulation equipment as the orthogonal systems employing correlation detection, requires only about twice as much received energy to achieve the same results for $L = 64$ quantization levels.

PAM appears to be better than PDM and PSK by a factor of three among the modulation methods which may be demodulated by linear synchronous detection, only because the average power is the criterion. If a peak power criterion were used, the three systems would behave equally well.

PAM and PSK occupy the least bandwidth of all the systems considered, $1/T$ cps, where T is the sample transmission time. The orthogonal systems, quantized PPM, FSK and PCM with orthogonal codes, occupy L/T cps, where L is the number of quantization levels. Nonredundant PCM occupies only $\log L/T$ cps. All these systems with the exception of PSK occupy half as much bandwidth if the adjacent channel carriers can be made phase coherent. Because it is impossible to make adjacent channels of PDM orthogonal to one another by appropriate frequency separation, PDM utilizes an excessive amount of bandwidth (approx $50/T$ cps) in order to maintain the channel cross-modulation less than 1%.

References

1. Shannon, C. E., and Weaver, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.
2. Lorens, C. S., and Viterbi, A. J., *Optimum Filtering*, External Publication No. 633, May 15, 1959.
3. Black, H. S., *Modulation Theory*, Van Nostrand Co., New York, 1953.
4. Woodward, P. M., and Davies, I. L., "Information Theory and Inverse Probability in Telecommunications," *Proceedings of the IEE*, 99 Part III: 37, 1952.
5. Viterbi, A. J., *On Coded Phase-Coherent Communications*, Technical Release No. 32-25, Aug. 15, 1960; to be published in IRE Transactions on Space Electronics and Telemetry, March 1961.

Table 1. Bandwidth occupancy for various detection and telemetry systems

Detection of modulation	Telemetry system	Bandwidth, cps: Adjacent channels not phase coherent[a]	Bandwidth, cps: Adjacent channels phase coherent[a]
Linear synchronous detection	PAM	$1/T$	$1/2T$
	PSK	$1/T$	$1/T$
	PDM	$50/T$, for less than 1% channel cross-modulation	
Correlation detection	Quantized PPM	L/T	$L/2T$
	Quantized FSK	L/T	$L/2T$
	PCM, orthogonal codes	L/T	$L/2T$
	PCM, nonredundant codes	$\log L/T$	$\log L/2T$

[a] The symbol T is the transmission time and L is the number of quantization levels.

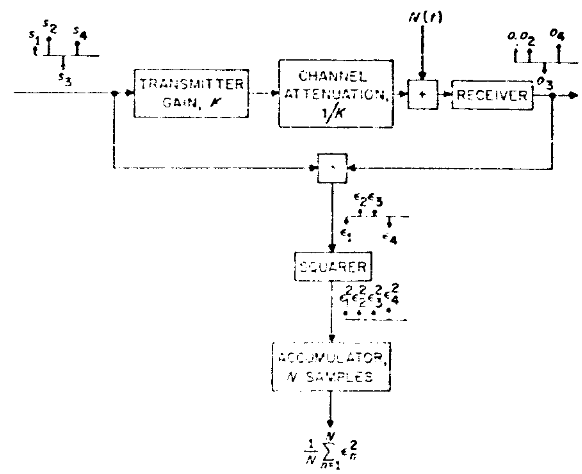

Fig. 1. Measurement of mean-square error

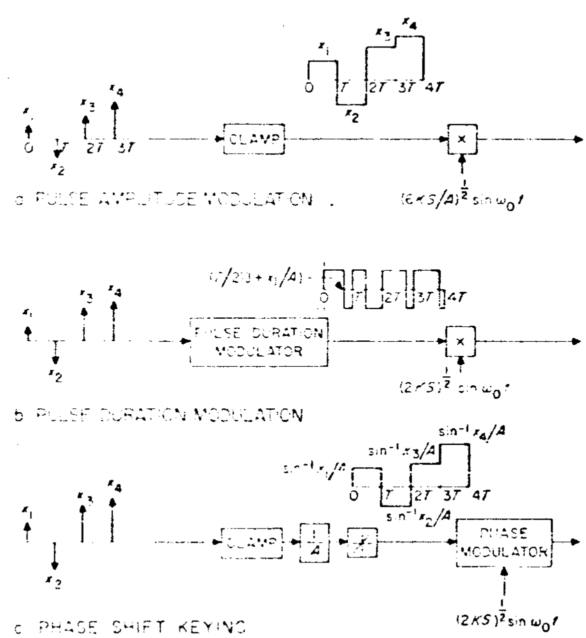

Fig. 2. Modulation systems pertinent to linear synchronous detection

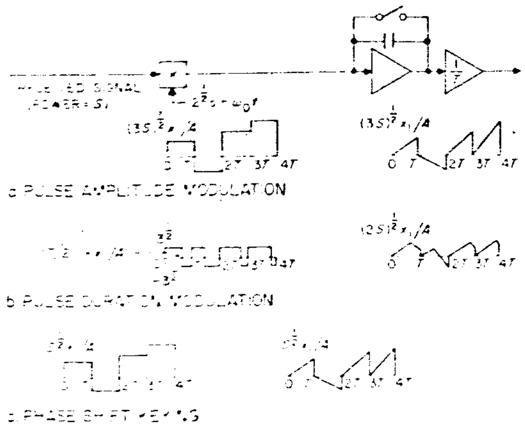

Fig. 3. Waveforms at input and output of pulsed integrator for linear synchronous detection

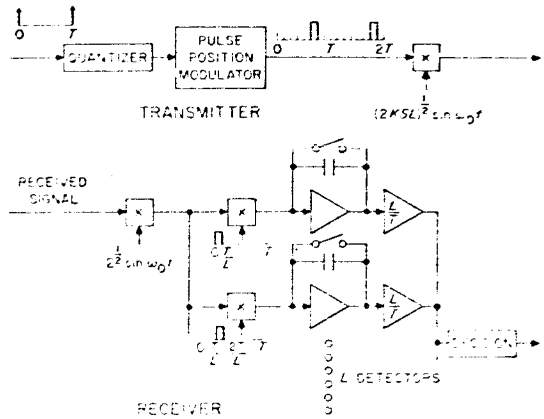

Fig. 5. Quantized pulse position modulation system employing correlation detection

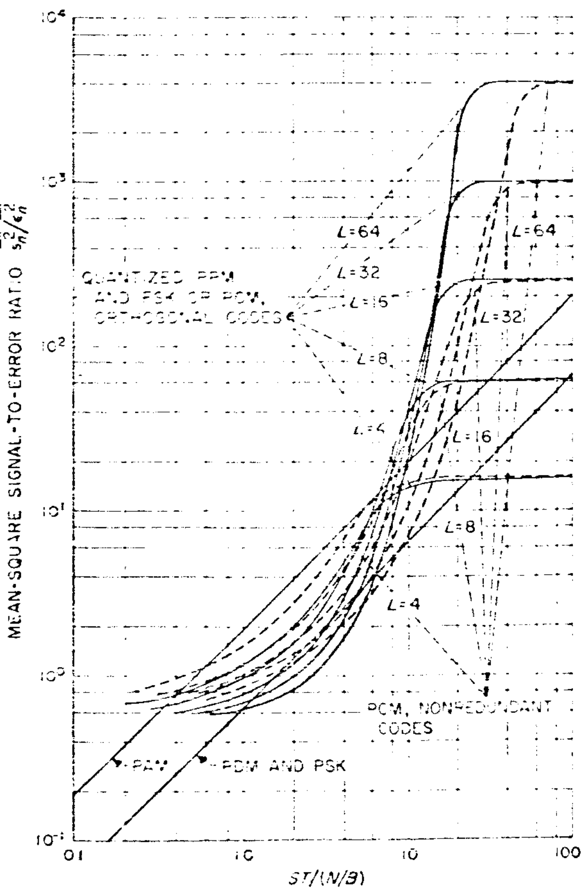

Fig. 4. Mean-square signal-to-error ratio

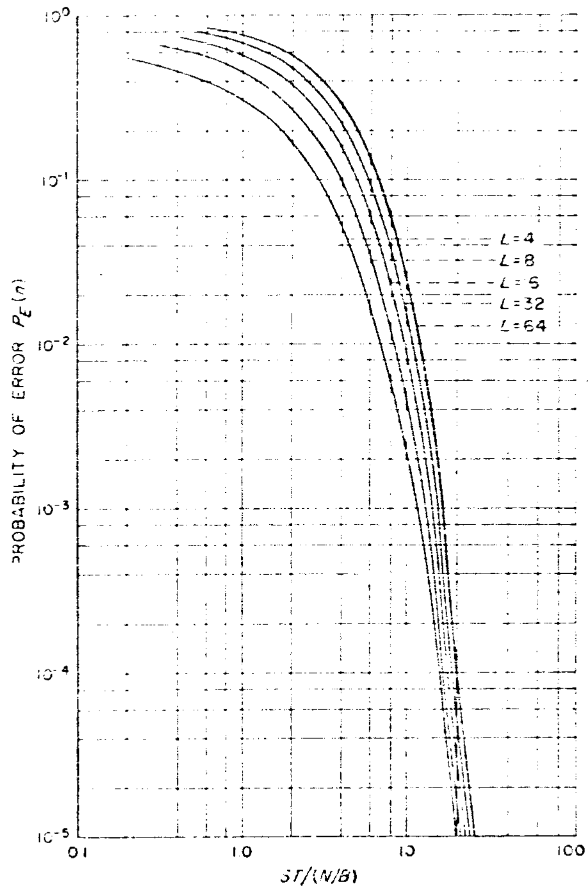

Fig. 6. Detection error probabilities for quantized PPM and FSK

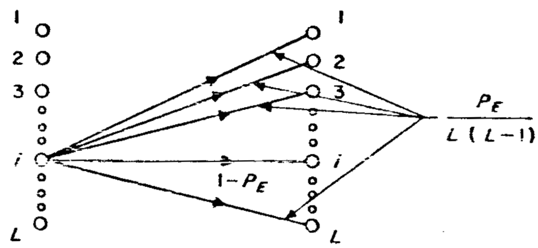

Fig. 7. Transition diagram for quantized PPM

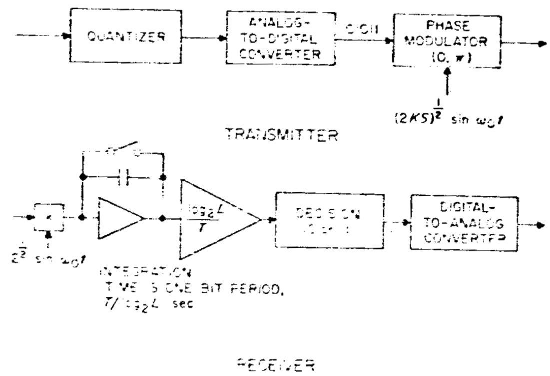

Fig. 10. Nonredundant PCM system

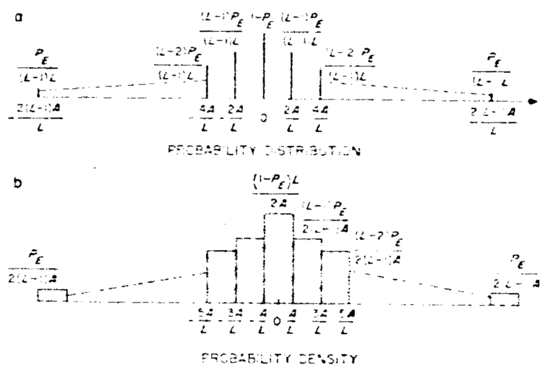

Fig. 8. Probability distribution and density of errors for quantized PPM

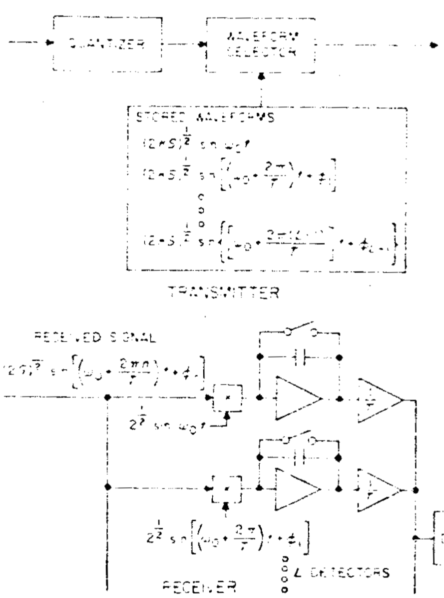

Fig. 9. Quantized FSK system employing correlation detection

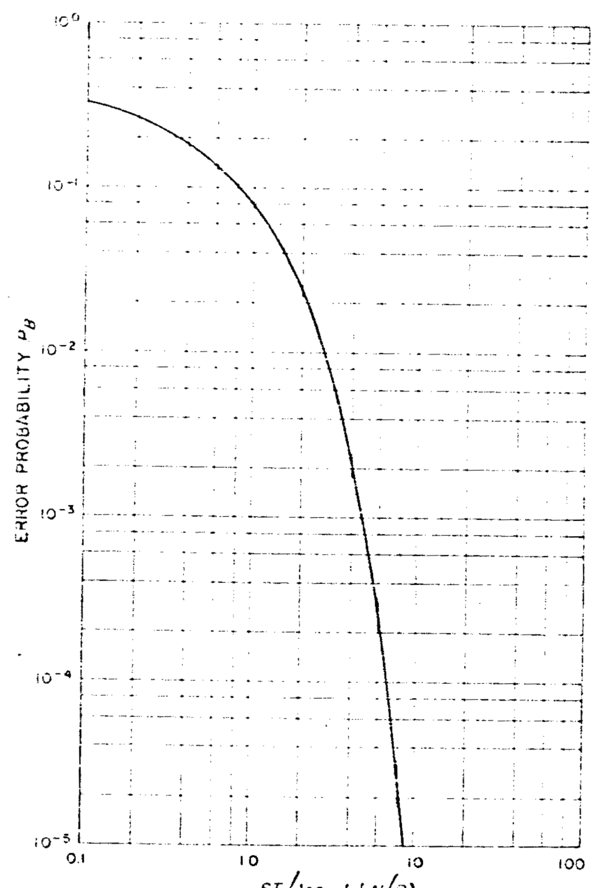

Fig. 11. Bit detection error probability for PCM, nonredundant codes

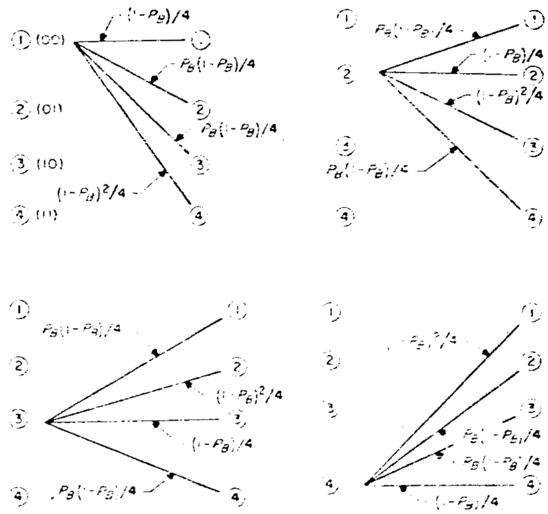

Fig. 12. Transition diagram for nonredundant PCM, four levels

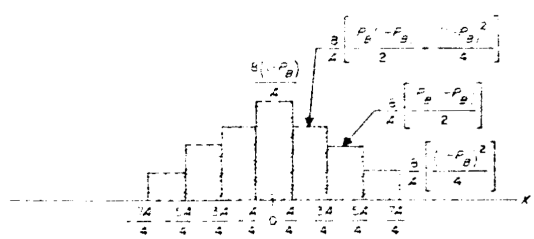

Fig. 13. Probability density of errors for nonredundant PCM, four levels